

# 尤度推定によるアクティブラーニング型授業の可視化

阪口 真也人<sup>1</sup> 豊浦 正広<sup>1</sup> 茅 暁陽<sup>1</sup> 埴 雅典<sup>1</sup>

概要：アクティブラーニング型授業では学習者の能動的な参加を促す必要がある。講師が授業を振り返るためには、授業がどのように推移したかを把握することが有用である。我々はこれまでに授業を収録した映像から、各時刻がどの授業カテゴリに分類される活動を行っているかの推定を行ってきた。本研究では、各時刻で何が行われたかを可視化することを目指す。授業カテゴリ認識の途中結果から得られる尤度を計算して表現することによって、カテゴリ分類の単純な認識結果の表示だけでは捉えられなかった授業の様子や特徴を可視化する。

## 1. はじめに

多くの教育現場においてアクティブラーニングが導入されてきている。アクティブラーニング型授業は講師が一方向的に知識伝達を行わず、学生が能動的に学習することにより、認知的、倫理的、社会的能力、教養、知識、経験を含めた汎用的能力の育成を図ることを目的としている。アクティブラーニング型授業には課題研究やPBL(Problem Based Learning, Project Based Learning)、ディスカッション、プレゼンテーションなどが積極的に取り入れられる。アクティブラーニング型授業は発展途上であり、授業での実施に関しては、経験的または理論的な知識が共有されてきてはいるものの、多くの講師にとってはその授業スタイルが確立してはいない。また、議題の提案や学生の誘導といった従来形式の授業では必要なかった指導法が講師に要求されるが、講師自身が授業の成否を客観的に認識することは容易ではない。授業の振り返り方法として、収録された授業映像の視聴が有効であるとされているが、長時間の動画視聴は講師に負担を強いることになり、授業へのフィードバックまでのサイクルを短期間で回すのは難しい。これらのことからアクティブラーニング型授業の客観的で即時的な分析が求められている。本研究ではアクティブラーニングの分析手法として収録された授業映像の自動解析と可視化を目指す。

我々は機械学習の一種である Bag-of- Words を用いてアクティブラーニング型授業状況の自動推定を行ってきた [1]。収録されたアクティブラーニング型授業の映像から、各時刻において何が行われているかを一意に推定し、

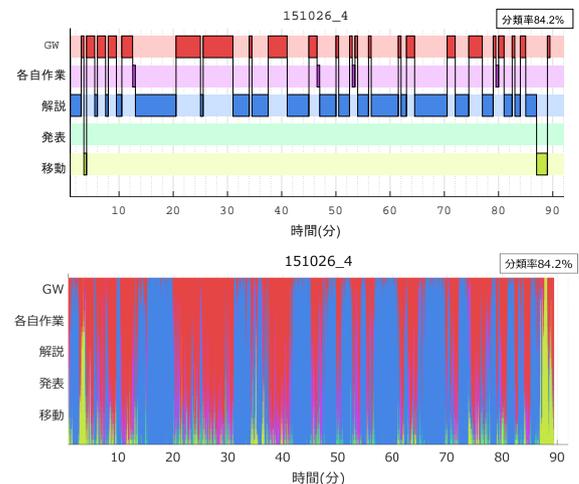


図 1 授業状況の可視化

Fig. 1 Visualization of class process

図 1 上のような表現による可視化を行った。

本研究ではこれまでの可視化に加えて、図 1 下に示すような尤度表現による可視化を行う。各時刻における教室内の状況を尤度という表現を用いて可視化を行うことで、一意の認識では確認できなかった授業カテゴリ間の差異や講師ごとの特徴をより捉えられることができると予想される。またこれらの特徴はアクティブラーニング型授業を行う講師にとっての新たな知見になることが期待される。

## 2. アクティブラーニング型授業のカテゴリ認識

既存手法のアクティブラーニング型授業映像のカテゴリ認識についてこの章で述べる。授業映像は大学で開講されたものを収録し、図 2 に示すように、収録が実施された教室には授業の人数やグループの数に応じて使用される

<sup>1</sup> 山梨大学  
University of Yamanashi



図 2 授業映像の一部

Fig. 2 An example shot of lesson video

稼働式の椅子および稼働式ホワイトボードが用意されている。また教室内の四方に向かってプロジェクタが設置されており、授業によって講師の立つ位置は異なる。授業の分類には、実際にアクティブラーニング型授業において頻度の高い5つの授業カテゴリを以下のように設定した。

- (1) 移動
- (2) 発表
- (3) 解説
- (4) 各自作業
- (5) グループワーク (GW)

(1) は授業間の移動や資料の配布, (2) は意見の発表や講師の質疑に対する応答, (3) は講師による学習指示や問題解説, (4) は小テストや問題演習等の個人での作業, (5) はグループワークや複数人での作業が含まれる。これらの授業カテゴリはいくつかの授業を観察し、漏れなく重なりがなくなるように設定した。

### 2.1 授業推定のためのデータ構築

映像から授業の展開を自動で認識するためのデータセットを構築する。自動推定には、各カテゴリの区別を可能にする特徴の設定が必要となる。そこで授業の様子を判断する特徴として、教室内の音量と動きの2つに着目した。教室内の音量は授業の盛り上がりと相関があると考えられ、議論中は教室内が賑やかであったり、問題の解答は静かに行われるなど、それぞれの授業カテゴリで異なる特徴があることが期待できる。また講師の話す音声においても、授業スタイルに応じて長さやタイミング等に違いが見られるなど、重要な特徴であることが期待される。教室内の動きの量は授業の活発度を表すと考えられ、手を動かして作業をするときと講師の話すときとは大きな差があることが期待される。これらから環境音、マイク音、フレーム間差分を特徴とする。教室内の音量(環境音)を収録を行うカメラから取得する。講師の音声(マイク音)はハンディマイクを用いて収録し、簡単なノイズ除去を行うことで音声入力の有無を明確に分離できるようにする。学生及び講師の動作量は授業映像のフレーム間差分から得られる差分を用いる。これらに加え、映像内のどこに動きが多く起

こったかという情報を、格子状に分割した映像のフレーム間差分から取得する。差分値の分布割合をヒストグラムで表し、4つ目の特徴として設定する。上記の4つの特徴は1秒おきにサンプリングを行うが、授業の状況は時間で推移するため、一時刻の特徴のみで判断することできない。そこで一時刻の特徴に加えて、数秒前からの時刻の特徴も考慮して一つの特徴群として設計する。

### 2.2 授業状況の推定

推定には機械学習の一種である Bag-of-Words を用いる。処理の概要を図3に示す。特徴の分類を行うために、学習データとして用いる映像のすべての時刻における特徴を特徴空間にプロットする。図3の例では視覚的に見やすいため、特徴空間を3次元で示す。特徴空間上の特徴点をk平均法(k-means法)によって類似しているものにクラスタリングをする。本研究では15のクラスタに分類する。次に任意時間の特徴の分布傾向とその時間で行われている授業カテゴリとの相関を学習するために、学習データの各時刻における特徴群のみを特徴空間にプロットする。特徴空間上の各ノードを、距離が最も近いクラスタに割り当て、ヒストグラムを作成する。これによって各時刻の特徴群からどのようなヒストグラムを得られるのかを学習する。またその時刻はどのような授業カテゴリに属するかという正解ラベルを与える。正解ラベルはあらかじめ目視によって映像から作成する。これにより各授業カテゴリのヒストグラムの形の傾向などを学習することが可能である。また学習に使用する映像のすべての時刻でのヒストグラムが学習データとして蓄積される。分類段階では初めに、推定を行う映像の任意時刻の特徴群から同様にヒストグラムを作成し学習データと比較をすることで授業カテゴリを認識することが可能である。ここでの識別にはk-NNを用いる。分類結果において、一つのカテゴリと判定される連続時間が

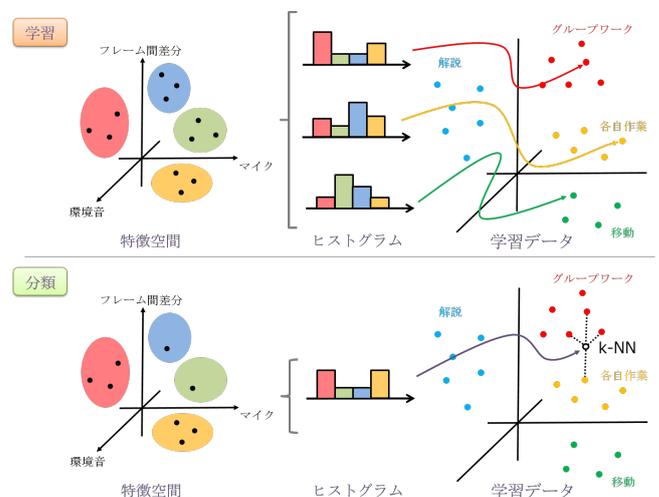


図 3 Bag-of-Words による認識

Fig. 3 Recognition by Bag-of-Words

表 1 自動分類の推定精度：講師 A

Table 1 Accuracy of automatic classification: Instructor A

| 授業 1  | 授業 2  | 授業 3  | 授業 4  | 授業 5  | 平均    |
|-------|-------|-------|-------|-------|-------|
| 90.3% | 61.9% | 84.2% | 47.3% | 79.5% | 75.4% |
| 授業 6  | 授業 7  | 授業 8  | 授業 9  |       |       |
| 84.9% | 78.6% | 72.7% | 79.4% |       |       |

表 2 自動分類の推定精度：講師 B

Table 2 Accuracy of automatic classification: Instructor B

| 授業 1  | 授業 2  | 授業 3  | 授業 4  | 授業 5  | 平均    |
|-------|-------|-------|-------|-------|-------|
| 70.1% | 80.0% | 49.2% | 90.6% | 66.1% | 72.5% |
| 授業 6  | 授業 7  | 授業 8  | 授業 9  |       |       |
| 56.5% | 58.6% | 85.1% | 97.1% |       |       |

極端に短く分類されてしまうことがあるが、授業内において一つの授業カテゴリが数秒ごとに切り替わるといことは想定されないため、1つの授業のカテゴリを30秒で区切り、この区間のカテゴリを対象時刻のカテゴリの多数決によって決定する。これによって分類されたカテゴリが断片化することを防ぐ。

### 2.3 授業展開の可視化

機械学習によって分類した授業展開の様子はタイムラインによる可視化を行う [2][3]。可視化の例を図 1 に示す。横軸が時間の流れ、縦軸が授業カテゴリを表す。授業の時系列における状況変化を表し、授業の流れや時間配分が一目で確認することができる。これにより授業時間と同じ長さの映像を、1枚の画像に集約して表現したことになる。また他講師や自身の他の授業と比較することで、より詳しく自身の授業展開の特徴を把握することができる。

複数のプロセスがあるときに、プロセス相互の関係を示すためにタイムラインと呼ばれる可視化手法が用いられる [2][3]。UML のシーケンス図や並列処理パイプラインでも同様の表現がよく見られる。本研究でも、それぞれの授業状況がどれくらいの時間行われているのか、あるいは、状況間の推移がわかりやすくなるように、5つの授業状況を並列化して可視化を行う。

### 2.4 認識精度

自動推定精度の評価実験を行う。講師 2 人の 90 分の授業映像をそれぞれ 9 本ずつ使い、Leave-one-out 交差検証 (LOOCV) を行った。Bag-of-Words によって、認識する映像の各時間における授業カテゴリがラベルとして得られる。これに対し、あらかじめ手作業によって作成した正解ラベルと比較することで推定精度を求める。

結果を表 1, 2 と可視化例を図 4 に示す。図 4(a) の例のように、平均値を上回る分類結果では、カテゴリの切り替わりのタイミングや経過時間を確認することが出来る。また図 4(b) の例に示すような平均値に近いものにおいても、カ

テゴリの誤分類はあるものの、授業状況の変化が起こった時間を予想することができる。図 4(c) と図 4(d) は講師 2 人の精度が低かった結果の例である。これらの結果では、グループワークのカテゴリでの誤分類が多く、各自作業や解説のカテゴリと認識している箇所が多く見られる。図 5 は学習に用いたデータの各カテゴリの割合を表している。このことから各自作業カテゴリは学習データが少なかったことが誤分類の原因の一つであることが考えられる。また他のカテゴリとの明確な差を判断する特徴が設定できていないことも要因の一つであることが予想される。

### 3. 尤度表現

ここでさらに授業状況の認識を行う際、各カテゴリが他とどの程度明確な差を付けて認識しているかという点に着目したい。図 4(a) と図 4(b) に示す結果では、グループワークと解説、グループワークと各自作業が細切れになって認識される例が見受けられた。図 4(a) の結果ではおおよそ正しく推定されているが、図 4(b) では各自作業でない時間帯が各自作業であると認識されてしまった。認識で何が起きているかを探るために、本研究ではさらに詳細な認識結果の可視化手法を提案し、問題の解析にあたる。アクティブラーニング型映像では、授業を行う講師ごとにその授業スタイルに個人差があるため、それらの特徴から認識された分類の誤りにも差があることが予想される。尤度による表現によって各講師の差をより明確にすることで、アクティブラーニング型授業スタイル共有のための知見が得られると期待できる。

一般物体認識の論文などで、スコアの順に認識結果を上位順に並べ、それぞれのスコアを合わせて表示する方法が用いられることがある [4][5]。1枚の画像や1本の映像の認識結果を提示するためには有用であるが、各時刻での認識結果を示すには十分ではない。対象となる各時刻での認識結果を示すために、我々は、対象時間全体でこの認識結果が俯瞰できるような可視化を提案する。また、特定の部屋や人に着目して、映像全体の様子を可視化する研究もある [6][7]。映像全体を一覧することはでき、それぞれの行動の尤度も表現されているものの、どの状況であるかを他の状況と比べて相対的に判断することは難しい。時間軸で変わる状況の推移を理解するためにも、表現方法としては十分ではない。

#### 3.1 投票結果による尤度算出

2.2 での自動推定の過程で行われる k-NN から、各カテゴリの尤度情報の取得を行い、新たな可視化手法を提案する。図 3 に示す分類段階において行われる k-NN では、新たに認識した時刻でのヒストグラムを学習データと比較し、似た特徴を多く持つノードのカテゴリに分類される。このときの周囲には似た特徴のノードが存在しているため、これ

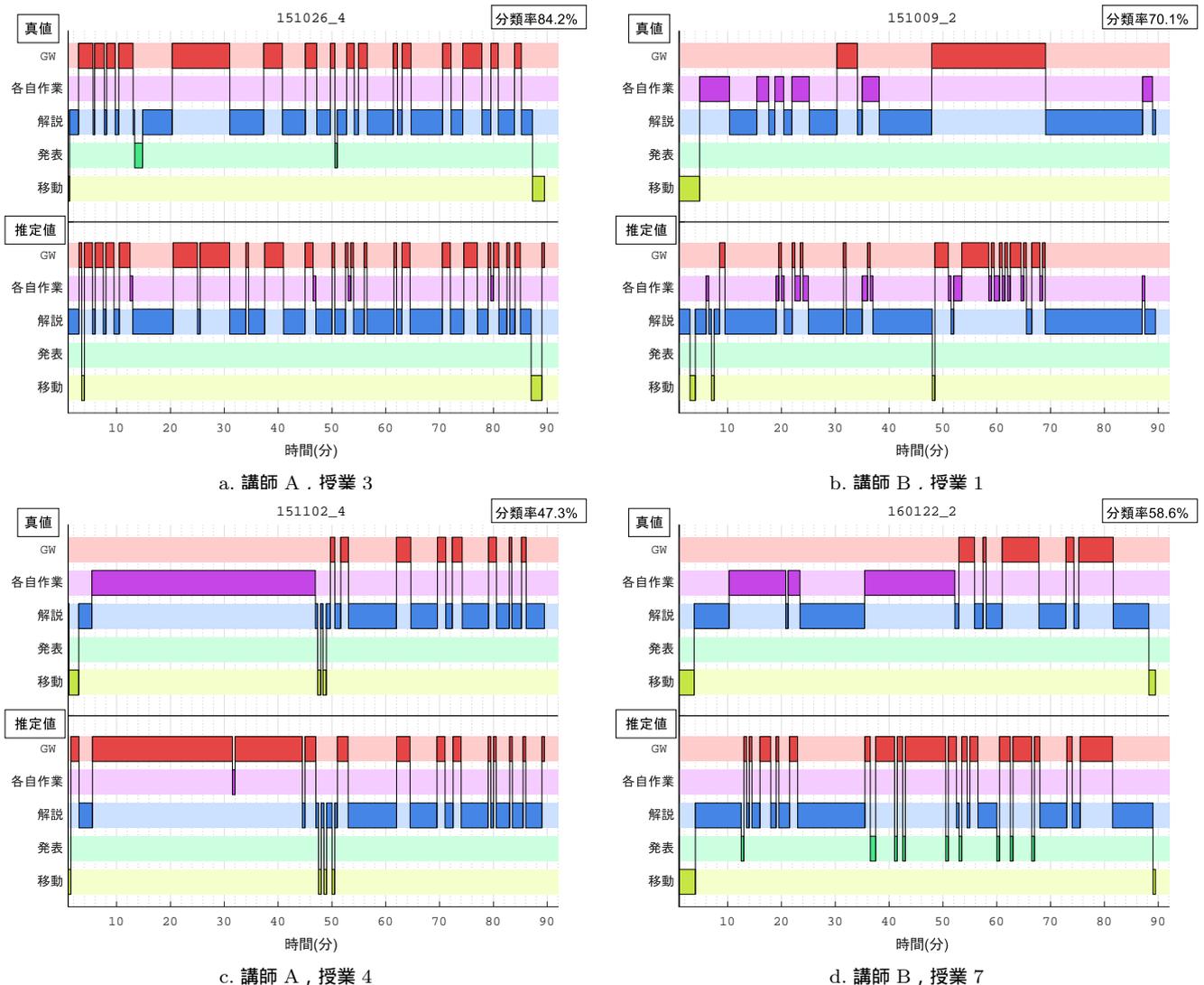


図 4 例

Fig. 4 An example of visualization of category transition

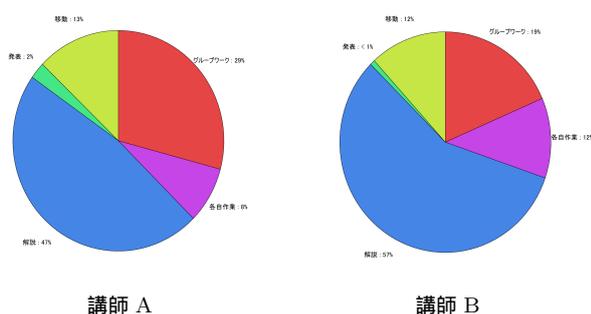


図 5 学習データの詳細  
Fig. 5 Learning data details

らのノードを複数確認することで得られる情報は、任意カテゴリにどの程度近いかという尤度を示すことが期待される。図 6 に本研究における尤度の定義を示す。時刻  $t$ 、状況  $l$  ( $l = 1, \dots, 5$ )、総投票数  $k$ 、各状況の  $k$ -NN 要素数  $s_t^l$ 、時刻  $t$  が状況  $l$  である尤度  $p_t(l)$  を式 (1) と定義する。従来の可視化手法では、推定される状況  $\tilde{l}_t$  を式 (2) で定義して

いた。このときには、 $p_t(l)$  の詳細が失われ、他のカテゴリである尤もらしさが表現できない。提案す手法は  $p_t(l)$  を可視化するものである。

$$p_t(l) = \frac{s_t^l}{k}, \sum_l \frac{s_t^l}{k} = 1 \quad (1)$$

$$\tilde{l}_t = \arg \max_l p_t(l) \quad (2)$$

尤度情報を得るために  $k = 31$  として、 $k$ -NN を行う。得られた尤度情報を図 6 に示すように投票数を各時刻で積み上げて可視化を行う。横軸が時間の流れ、縦軸が任意時間における類似カテゴリを表し、下段は  $k$ -NN の投票結果を示す。各列に対して、類似カテゴリの量を視覚的にわかりやすくするために状況でソートを行っている。この手法によって尤度を可視化する。

### 3.2 考察

可視結果を図 7 に示す。図 7(a) の講師 A による授業は、

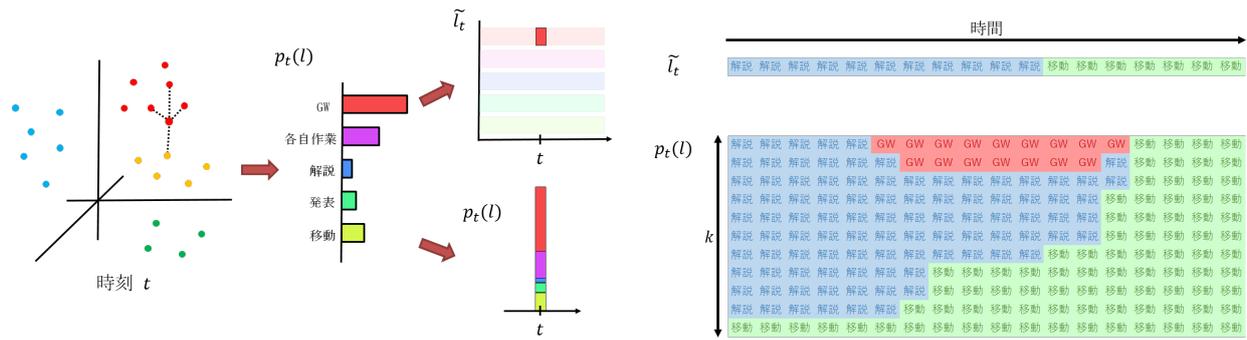


図 6 尤度の定義

Fig. 6 Definition of likelihood of individual states

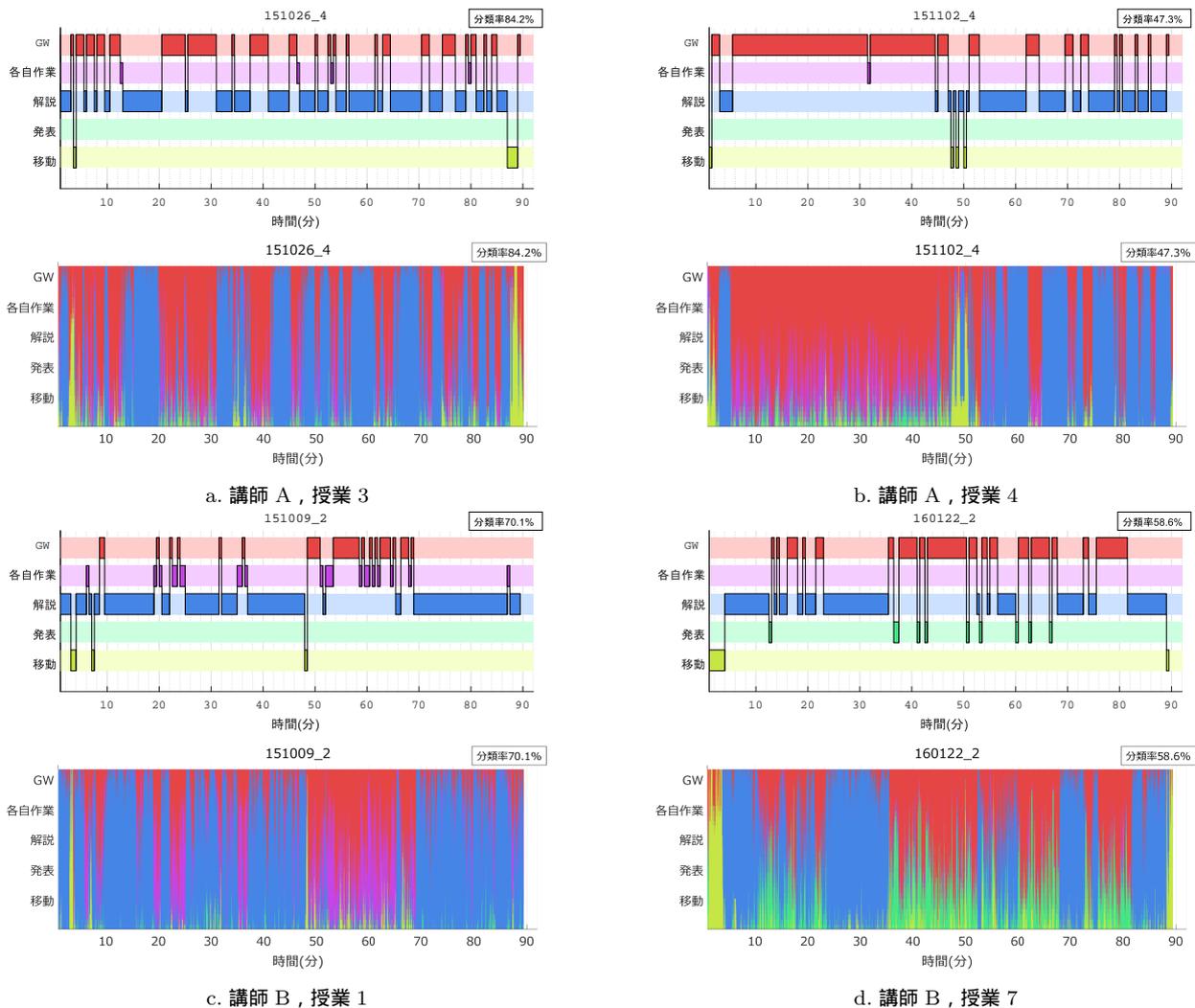


図 7 尤度可視化

Fig. 7 Likelihoody visualization

図 7(c) の講師 B の授業に比べて色がはっきりと分かっていることから、曖昧さが少ないことがわかる。授業の活動状況が他の活動と明確に分かれていることでこのような結果になったことが考えられる。またこの曖昧さが分類精度に大きく影響を及ぼしていることも見て取れる。分類精度の高い授業は低いものと比べて曖昧さが少ない。授業を行う上で、アクティブラーニングの活動を明確に行えているか

どうかは、授業改善に重要な点であることが予想される。図 7(b) と図 7(d) は 2.4 で誤分類の多かった 2 授業である。それぞれ誤分類箇所において曖昧が多いことがわかる。また講師 A (図 7(b)) は、グループワークと各自作業のカテゴリの曖昧さが大きいのに比べ、講師 B (図 7(d)) はグループワークと発表のカテゴリに曖昧さが大きいことが分かった。この曖昧さを減らすことは、自動での授業の状況推定



図 8 曖昧さを含む授業映像の例

Fig. 8 An example shot of lesson video including ambiguity

精度を高めるのに加え、講師ごとの特徴が反映されることから、授業改善に繋がる可能性があると考えます。本研究ではアクティブラーニング型授業で実際に行われた授業状況を、頻度の高い5カテゴリに絞って分類を行ったが、これら5つのカテゴリ以外の新たなカテゴリの設定が必要であるのではないかと考える。

図8はグループワークと解説のカテゴリの曖昧さが多く見られた時刻での例である。教室全体ではグループワークが行われているが、講師とグループとの対話によって解説カテゴリに分類されている。このような特定の時刻で2状態の投票数が拮抗している状況を映像から認識することは難しく、一意に授業カテゴリを決定することは容易ではない。尤度の可視化はこのような曖昧さを含む状況の検出を可能にした。

#### 4. おわりに

収録した授業映像から自動で授業状況を推定し、尤度表現による可視化を行うことで、授業状況の容易な振り返りを可能にし、講師ごとの特徴を確認することができた。アクティブラーニング型授業の分類を行う際の手数による真値データの作成においても、その曖昧さから個人差は生まれてしまうことがこれまでにわかっている。講師が自身で行っている授業内容と客観的に判断された授業内容の差を示すことで、効果的な振り返りができると期待される。

謝辞 本研究の一部は、JSPS 科研費 JP16K12784, JP26282062, JP15K00499 の助成を受けたものである。

#### 参考文献

- [1] M. Toyoura, M. Sakaguchi, X. Mao, M. Hanawa, Masayuki Murakami, “Visualizing the Lesson Process in Active Learning Classes,” IEEE FIE, Article 4, 2016.
- [2] C. Plaisant, B. Milash, A. Rose, S. Widoff, B. Shneiderman, “LifeLines: visualizing personal histories,” Proc. SIGCHI conference on Human factors in computing systems, pp. 221-227, 1996.
- [3] R. Miyazaki, I. Fujishiro, R. Hiraga, “comp-i: A system for visual exploration and editing of midi datasets,” 2004.
- [4] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Suk-

thankar, L. Fei-Fei, “Large-scale video classification with convolutional neural networks,” Proc. IEEE Computer Vision and Pattern Recognition, pp. 1725-1732, 2014.

- [5] A. Krizhevsky, I. Sutskever, G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” Advances in neural information processing systems, pp. 1097-1105, 2012.
- [6] M. Romero, J. Summet, J. Stasko, G. Abowd, “Viz-A-Vis: Toward visualizing video through computer vision,” IEEE Transactions on Visualization and Computer Graphics, vol. 14, pp. 1261-1268, 2008.
- [7] M. Toyoura, S. Nishiguchi, X. Mao, and M. Murakami, “ActVis: Activity visualization in videos,” Proc. Cyberworlds, pp. 222-225, 2013.